

Deep Learning for Computer Vision 2 assignment

Lorenzo Zanolin

Results

Model	mAP
FASTERRCNN_MOBILENET_V3_LARGE_320_FPN	0.5865
FASTERRCNN_MOBILENET_V3_LARGE_FPN	0.6935
FASTERRCNN_RESNET50_FPN	0.7535
FASTERRCNN_RESNET50_FPN_V2	0.6745

Comments

Our dataset consisted of 4 classes: *cars*, *furnitures*, *moto*, *fruits*. We used the VGG Image Annotator to manually detect the objects in our dataset images; CoCo Format was the one selected for the annotations. Some examples can be found in the dataset folder. All the tests were done using Faster R-CNN model, with different backbones:

- MobileNetV3-Large FPN.
- MobileNetV3-Large backbone tuned for mobile use cases.
- ResNet-50-FPN backbone from paper [2].
- ResNet-50-FPN backbone from paper [1].

We trained for 5 epochs, using SGD as optimizer and a Learning rate of 0.005. As we can expect, F-RCNN with MobileNetV3-Large FPN performed slightly better than the one with backbone tuned for mobile use cases since it has a lower resolution, according to PyTorch official documentation. The best results were obtained by the RESNET50 model by a little increase of performance w.r.t. the other models, despite the amount of time spent (about an hour and a half).

References

- [1] Yanghao Li, Saining Xie, Xinlei Chen, Piotr Dollar, Kaiming He, and Ross Girshick. Benchmarking detection transfer learning with vision transformers. *arXiv preprint arXiv:2111.11429*, 2021.
- [2] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.